

基于贝叶斯网络的民航突发事件因果关系分析方法研究 *

王 红, 郭笑丹, 祝 寒

(中国民航大学 计算机科学与技术学院, 天津 300300)

摘 要: 针对民航突发事件因果关系无法有效评估与关联分析的问题, 提出了一种基于贝叶斯网络的民航突发事件因果关系分析方法。在民航突发事件应急管理领域本体的基础上引入贝叶斯理论, 首先通过规则设计实现了领域本体中概念、关系与实例的贝叶斯网络转换, 然后采用贝叶斯网络知识合成算法 E-IPFP 构建贝叶斯网络节点的条件概率表, 并通过消息传递机制计算父子节点间的概率关系, 获得民航突发事件因果关系的概率分布。采用民航突发事件应急管理领域本体和世界民航事故调查跟踪报告中的案例作为实验数据, 给出了民航突发事件因果间关系的分析, 为基于大数据的突发事件关联分析与推理提供了方法支持。

关键词: 民航突发事件; 领域本体; 贝叶斯网络; 因果关系; 条件概率

中图分类号: TP399 **doi:** 10.3969/j.issn.1001-3695.2017.09.0926

Research on causal reasoning method of domain ontology based on Bayesian network

Wang Hong, Guo Xiaodan, Zhu Han

(School of Computer Science & Technology, Civil Aviation University of China, Tianjin 300300, China)

Abstract: According to the fact that civil aviation emergency cannot effectively evaluate causality and correlation analysis, this paper proposed a causal analysis method for civil aviation emergencies based on Bayesian Network. The method introduces Bayesian theory on the basis of the domain ontology of civil aviation emergency management. Firstly, the Bayesian network realizes the transformation of concepts, relations and instances in domain ontology with rules design. Then construct the conditional probability table for Bayesian network nodes using Bayesian network knowledge synthesis algorithm E-IPFP, and calculate the probability relationship between parent nodes and child nodes through the message passing mechanism, to obtain the probability distribution of civil aviation incidents causation. This paper adopts the civil aviation emergency management domain ontology and the world civil aviation accident investigation tracking report as experimental data, gives the analysis of the causal relationship between civil aviation emergencies, which provides the method to support the correlation analysis and reasoning of unexpected events based on large data.

Key Words: civil aviation emergencies; domain ontology; bayesian networks; causality; conditional probability

0 引言

民航突发事件应急管理是为了应对民航突发事件而进行的一系列有计划有组织的管理过程, 主要任务是如何有效地预防和处置各种突发事件, 最大限度地减少突发事件的负面效应、人员伤亡和财产损失[1]。事件间因果关系[2-3]的分析对民航突发事件应急管理中的预防与处置具有重要的意义。目前基于领域本体的民航突发事件应急管理的相关研究主要集中在领域本体的构建、采用 SWRL 和描述逻辑等基于规则的应急方案的推理和相似案例的检索[4,5,6]等方面, 缺乏基于领域本体的突发事件因果关系分析与推理的研究。

贝叶斯网络^[7]BN(Bayesian network)于1985年由 Judea Pearl 首先提出, 是一种基于概率推理的图形化网络, 针对不确定性知识推理与数据分析的模型^[8,9,10]。蔡炳万^[11]等提出了一种基于本体的贝叶斯网络知识推理方法, 主要思想是依据产品设计过程中设计任务、设计人员、设计知识三个本体模型来建立贝叶斯网络的拓扑结构, 没有考虑本体与贝叶斯网络的转换问题。张什永^[12]等针对本体到贝叶斯网络转换中概率不一致的问题, 提出了一种改进的贝叶斯网络知识合成算法 E-IPFP^[13], 该方法通过加入约束集合有效的解决了 BN 中概率不一致时的合成问题。

本文将 E-IPFP 算法与民航突发事件应急管理领域本体相

基金项目: 国家自然科学基金资助项目 (U1633110)

作者简介: 王红 (1963-), 女, 重庆万州人, 教授, 主要研究方向为民航信息系统与应急管理、智能信息处理、数据挖掘; 郭笑丹 (1993-), 女 (通信作者), 硕士, 主要研究方向为智能信息处理、数据挖掘 (295449302@qq.com); 祝寒 (1992-), 女, 硕士研究生, 主要研究方向为智能信息处理、数据挖掘。[在此处键入]。

结合, 设计了一组将领域本体转换为贝叶斯网络的规则, 实现了本体与贝叶斯网络的转换; 在获取突发事件因果关系证据节点概率的基础上, 采用 E-IPFP 算法与消息传递机制进一步计算节点的后验概率, 从而获得突发事件的因果关系的概率分布关系, 该方法旨在解决基于领域本体的民航突发事件因果关系关联分析与推理问题, 为突发事件的预防与处置提供更好的数据支持。

1 研究思路

基于贝叶斯网络的民航突发事件因果关系分析思路如图 1 所示。

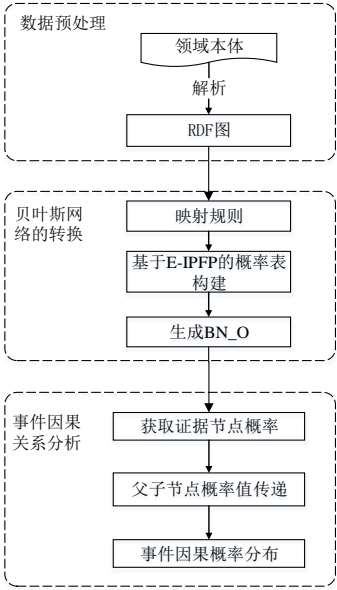


图 1 基于贝叶斯的民航突发事件因果关系分析过程

a)数据预处理。将民航突发事件应急管理领域本体解析为 RDF 图, 并根据世界民航事故调查跟踪信息报告^[14]中的专家分析对 RDF 图中的节点设置概率属性生成节点的概率文件。

b)贝叶斯网络的转换。设计规则实现领域本体 RDF 图到贝叶斯网络 BN 的转换, 并采用贝叶斯知识合成算法 E-IPFP 实现概率的合成, 生成领域本体贝叶斯网络 BN_O。

c)事件因果关系分析。采用基于消息传递机制的贝叶斯网络推理, 实现事件因果关系概率值传递, 获得事件因果关系的概率分布。

2 基于领域本体的 BN 构建方法

贝叶斯网络包括一个有向无环图 DAG(directed acyclic graph)和一个条件概率表集合 CPT(conditional probability table)。

定义 $BN=\{G, P\}$, G 代表有向无环图 $G=\{V, E\}$, 其中 V 表示

图中的节点(随机变量)集合, E 表示有向边集合代表变量之间的关系; P 表示节点条件概率表的集合, 每个节点对应一个条件概率分布表, 该表是对节点变量与其父节点之间关系程度描述。若存在节点 V_i 直接影响到节点 V_{i+1} , 即 $V_i \rightarrow V_{i+1}$, 则建立有

向弧 (V_i, V_{i+1}) , (V_i, V_{i+1}) 的权值(即连接强度)用条件概率 $P(V_{i+1}|V_i)$ 来表示, 存储于条件概率表中。

2.1 转换规则的定义

领域本体是对概念、概念间关系的描述, 包括类、关系、实例, 领域本体 RDF 是由一系列三元组(主体, 谓语, 客体)组成的, 其中主体、客体代表本体中的概念 C , 谓语代表本体中的关系 R , 这些三元组形成如图 2 所示的领域本体 RDF 图。

根据以下规则将领域本体 RDF 图转换为贝叶斯网络:

规则 1 对于本体 RDF 图中的概念集合 $C=\{C_1, C_2, \dots, C_n\}$, C_i 对应到贝叶斯网络中的节点 $VC_i=(a, \bar{a})$ ($i \in n$), (a, \bar{a}) 分别表示概念 C_i 的状态

属性中 True 和 False 的概率值。

规则 2 $R=\{R_1, R_2, \dots, R_S\}$ 表示本体的关系集合, 若本体中概念 C_i 与概念 C_{i+1} 存在关系 R_x , 且 C_i 直接影响 C_{i+1} , 则将两个概念之间的影响关系 R_x ($x \in S$) 映射为 BN 中节点 V_i 与 V_{i+1} 之间的有向边 E_x 。

规则 3 对于本体中实例 $I=\{I_1, I_2, \dots, I_m\}$, 将 I_j ($j \in m$) 对应到贝叶斯网络中的节点 $VI_j=(a, \bar{a})$, (a, \bar{a}) 分别表示实例 True 和 False 的概率值。

映射过程如图 2 所示。

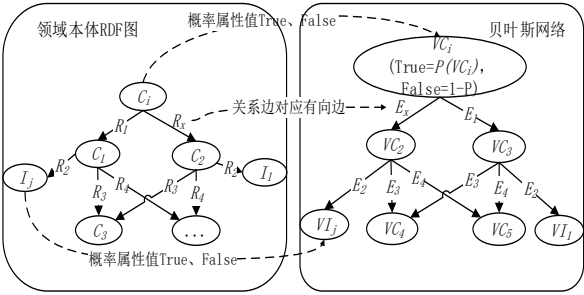


图 2 映射过程

2.2 基于 E-IPFP 的 BN 节点条件概率表构建

2.2.1 基于 E-IPFP 的贝叶斯知识合成算法

贝叶斯知识合成算法是将给定的概率知识合成到贝叶斯网络的过程, 基本思路如下:

a)输入联合概率 $P_0(X)=\prod_{i=1}^n P_0(X_i|\pi_i)$, 并将 BN 中概率

集合 $D=\{D_1, D_2, \dots, D_m\}$ 作为约束条件;

b)判断其新的约束条件下是否收敛, 若收敛, 计算当前联合概率分布 $P_k(X)$ ($k \in m$): 若 $P_{k-1}(Y^i)=0$, 则 $P_k(X)=0$, 否则

$$P_k(X) = P_{k-1}(X) \bullet \frac{D(Y^i)}{Q_{k-1}(Y^i)} \quad (1)$$

循环计算直至收敛;

c)若不收敛, 则分别在约束集与概率上加入 α 参数,

$D_k(Y^i) = (1-\alpha)D(Y^i) + P_{k-1}(Y^i)$ 作为新的约束循环计算 $P_k(X)$

直至收敛。

其中: X 代表贝叶斯网络中的节点随机变量, n 代表变量个数,

$P(X|\pi)$ 为节点的条件概率, $\alpha \in (0,1)$ 为约束条件的参数,

$P_k(X)$ 表示当前的联合分布, $Q(Y)$ 表示子节点联合概率分

布, Y 为 X 的子节点集, π 为 X 的父节点集合, D 表示新的约束条件, m 为约束集个数。

2.2.2 BN 节点条件概率表构建

节点条件概率表的构建是将 RDF 图中节点概率文件中 True 和 False 的状态值转换到 BN 中节点条件概率表的过程, 转换的方法:

a)本体 RDF 图中节点概率属性值不为空时, 即当节点 C_i 直接影响到节点 C_{i+1} 时, 则将其对应于 BN 中的 V_i 、 V_{i+1} 节点, 此时 C_{i+1} 的条件概率为 $P_{C_{i+1}}(a_{i+1}|a_i)$, 其中 a_i 代表

$C_i = \text{True}$, a_{i+1} 代表 $C_{i+1} = \text{True}$, 在概率文件中的表示为

```
<owl:Probability rdf:ID="P(ai+1|ai)">
<hasProposition>ai+1</hasProposition>
<hasCondition>ai</hasCondition>
<hasValue>0.63</hasValue>
</owl:Probability>
```

转换到贝叶斯网络中以条件概率表的形式存储, V_i 直接影响到 V_{i+1} 并作为 V_{i+1} 的父节点, 节点 V_{i+1} 的条件概率如表 1 所示。

表 1 V_{i+1} 节点条件概率

V_i	$P(V_{i+1} V_i)$
True	0.63
False	0.371

b)本体 RDF 图中节点概率属性值为空时, 则在转换过程中, 遵循默认值的设定, 即节点的状态初始值设为 $0.5(\text{True}=0.5, \text{False}=0.5)$ 。

设领域本体中事件的原因 S 、结果 R 、事件严重程度 G 为 3 个类概念, 事件实例集合 E , 根据映射规则, 采用以下的 E-IPFP 的贝叶斯知识合成算法进行转换:

(a)采用规则 1 为 RDF 图中概念节点 C_i 设置状态属性 True 和 False 并赋值, 其中突发事件类中事件原因节点 S_i 转换成 BN 中的节点 VS_i , 事件结果节点 R_i 转换成 BN 中的节点 VR_i , 事件严重程度节点 G_i 转换为 BN 中的节点 VG_i ;

(b)采用规则 2 对突发事件本体中事件实例、原因、结果之

间的关系 R_i 进行映射转换为网络中的有向边 E_i , 并严格遵守上下层次关系, 上层概念为下层概念的父节点;

(c)采用规则 3 对 RDF 图中事件实例节点 E_i 进行转换, 为节点状态属性赋值并对应到贝叶斯网络中的节点 VE_i , 根据其原因属性与结果属性决定父子节点; 根据人员伤亡、经济损失、航空器损坏程度 3 个数据属性值确定节点 VE_i 到事件严重程度节点 VG_i 的连接关系。

转换后的领域本体贝叶斯网络如图 3 所示。

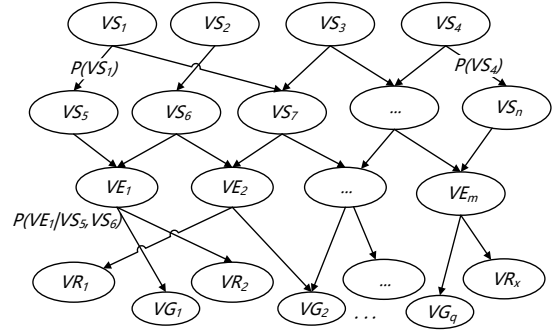


图 3 领域本体的贝叶斯网络(BN_O)

2.3 基于消息传递的贝叶斯推理

消息传递是根据节点集合 V 中每个节点自身的条件概率 $P(V)$ 与相邻节点传递概率值来求解后验概率^[15]的过程。定义网络中节点状态 $\text{True}=1$ 的节点为证据节点 V , 反之为非证据节点 V' , 如图 4 所示子节点 V'' 向父节点 V' 传递的消息称为 λ 消息, 父节点返回给子节点称为 ρ 消息, 通过给定的证据来计算所有节点的后验概率 $P(V')$ 。

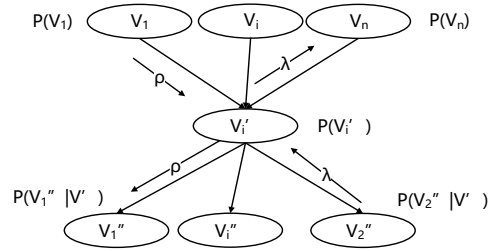


图 4 节点消息传递过程

基于消息传递的推理算法:

a) 输入条件概率 $P(V)$, 证据节点集合 $V = \{V_1, V_2, \dots, V_i, \dots, V_n\}$ 。

b)如果非证据节点 V' 有父节点, 则受到其父节点传递的消息, 用式(2)求出 $\rho_{V_i}(V'_i)$ 。

$$\rho_{V_i}(V'_i) = \sum_{V_i} P(V_i|V'_i) \prod_{j=1}^p \rho_{V_j}(V'_i) \quad (2)$$

如果 V' 有子节点, 子节点集合为 V'' , 用公式(3)计算 $\lambda_{V_i}(V'_i)$

$$\lambda_{V_i}(V'_i) = \prod_{j=1}^m \lambda_{V''_j|V'_i}(V'_i) \quad (3)$$

发事件 U_1 与非航空器突发事件 U_2 ; 与 U_1 、 U_2 事件相关的原

因 $S = \{S_1, S_2, \dots, S_i, \dots, S_n\} (i \in n)$, S_i 的子类为 S_{ij} ; 突发事

件中与航空器相关的事件结果如表 2 所示; 事件严重程度 G_i

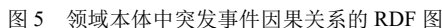
分为：一般 G_1 、较大 G_2 、重大 G_3 、特大 G_4 四个类别。

表2 民航突发事件中与航空器相关的事件结果

ID	事件结果
R_1	航空器坠毁
R_2	航空器失联
R_3	航空器重着陆
R_4	航空器暂停使用
R_5	航空器空中遇袭
R_6	航空器冲偏出跑道
R_7	跑道入侵

民航突发事件应急管理领域本体中的核心概念包含应急预案、应急资源、应急演练、突发事件、应急处置、资源报告、善后处置等类,突发事件及其因果关系的RDF图如图5所示。

其中:突发事件 $U=\{U_1, U_2, \dots, U_i, \dots, U_n\} (i \in n)$ 包含航空器突



```
<hasVariable>航空器突发事件</hasVariable>
<hasState>True</hasState>
</owl:Proposition>

<owl:Proposition rdf:ID="b1">
<hasVariable>航空器突发事件</hasVariable>
<hasState>False</hasState>
</owl:Proposition>

<owl:Probability rdf:ID="P(a1|a)">
<hasPropCollection>m1</hasPropCollection>
```

```
<owl:Variable rdf:ID="航空器突发事件">
<hasClass>CAEDO.航空器突发事件</hasClass>
</owl:Variable>
<owl:Proposition rdf:ID="a1">
```


<hasValue>0.5</hasValue>

</owl:Probability>

</owl:Proposition>

本体数据与概率数据作为输入, 采用 2.2.1 的方法将概率属性值 $P(C)$ 作为算法的输入, 形成完整的贝叶斯网络 BN_O。其中突发事件原因节点 S_i 对应 BN 的节点 VS_i ($i \in n$); 事件结果 R_i 对应为叶子节点 VR_i ($i \in x$); 事件严重程度 G_i 对应为叶子节点 VG_i ($i \in q$); 事件实例 E_i 对应贝叶斯网络中的节点 VE_i ($i \in m$)。

3.2 基于消息传递的贝叶斯推理

采用 2.3 算法设置 BN_O 中的证据节点集合, 以集合中节点条件概率作为消息传递推理算法的输入, 对节点的后验概率进行推理。

设节点 VE_1 、 VE_2 代表非航空器突发事件 (如: 7.17MH17 马航客机被导弹击中坠毁事件、8.24VD8387 河南航空 8 月 24 日伊春坠机事件), VE_3 、 VE_4 代表航空器突发事件 (如 11.4QF-32 澳大利亚快达航空飞机发动机非包容性失效事件、7.29MS-667 埃及航空飞机驾驶舱失火事件); 节点 VS_{i1} (飞行员操纵失误)、 VS_{71} (空管指挥不当)、 VS_{n2} (天气原因) 代表事件原因。

基于消息传递的推理实现如下:

输入: 贝叶斯网络 BN_O (网络结构和条件概率表集合),

证据节点 VE_2 取值状态 $\text{True}=1$;

输出: 非证据节点 VS_{i1} 、 VS_{71} 、 VS_{n2} 的条件概率分布: $P(VS_{i1})$ 、 $P(VS_{71})$ 、 $P(VS_{n2})$;

设置证据节点 VE_2 传递消息初始值:

$$\rho_{VE_2}(F) = 0.0, \quad \rho_{VE_2}(T) = 1.0$$

$$\lambda_{VE_2}(F) = 0.0, \quad \lambda_{VE_2}(T) = 1.0$$

根据非证据节点 VS_{i1} 、 VS_{71} 、 VS_{n2} 的条件概率表设置如下:

$$\begin{aligned} \rho_{VS_{i1}}(F) &= 0.45, & \rho_{VS_{i1}}(T) &= 0.55 \\ \rho_{VS_{71}}(F) &= 0.50, & \rho_{VS_{71}}(T) &= 0.50 \\ \rho_{VS_{n2}}(F) &= 0.40, & \rho_{VS_{n2}}(T) &= 0.60 \end{aligned}$$

通过式(3)计算出 VE_2 的 λ_{VE_2} :

$$\lambda_{VE_2}(F) = 0.25, \quad \lambda_{VE_2}(T) = 0.74$$

对于节点 VS_{i1} 、 VS_{71} 、 VS_{n2} , 获得子节点传递的消息 λ , 输入 ρ 、 λ 值采用公式(6)计算 $P(VS_{i1})$ 、 $P(VS_{71})$ 、 $P(VS_{n2})$; 同理计算出节点 VE_2 的后验概率。

直至整个网络中无消息传递, 输出节点的后验概率, 图 6 给出了与事件因果关系相关的部分概率分布计算结果。

3.3 效果分析

根据图 6 中领域本体突发事件因果关系的概率分布以及各个节点的条件概率表, 在以节点 VE_i 为证据节点的计算中, 可得出事件原因概率分布, 图中事件节点 VE_2 的条件概率表如表 3 所示。

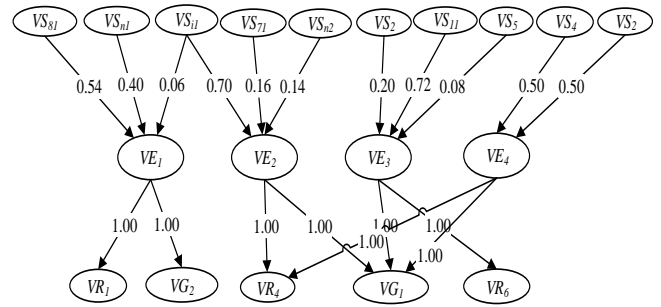


图 6 突发事件因果概率分布部分结果

表 3 VE_2 节点的条件概率

VS_{i1}	VS_{71}	VS_{n2}	$P(VE_2 VS_{i1}, VS_{71}, VS_{n2})$
T	T	T	1
T	T	F	0.86
T	F	T	0.84
T	F	F	0.7
F	T	T	0.3
F	T	F	0.16
F	F	T	0.14
F	F	F	0

表 3 给出了图 6 中节点 VE_2 代表的突发事件与 3 个原因父节点 VS_{i1} 、 VS_{71} 、 VS_{n2} 之间的的概率传递关系。其中:

a) 当 3 个原因节点 VS_{i1} 、 VS_{71} 、 VS_{n2} 状态都为 T (即 3 个原因都存在) 时, VE_2 事件发生的概率为 1。

b) 当 3 个原因节点 VS_{i1} 、 VS_{71} 、 VS_{n2} 中同时存在 2 个原因时, 事件发生的概率与原因 1 和原因 2 的相关程度会不一样, 表 3 中概率为 0.86 时与 VS_{n2} 更相关, 而概率为 0.84 时则与 VS_{71} 更相关, 特别值得注意的是概率为 0.3 时与 VS_{i1} 更相关, 反映出不同原因对事件的影响差异较大。

c) 当 3 个原因节点 VS_{i1} 、 VS_{71} 、 VS_{n2} 中只存在 1 个原因时, 不同原因导致事件发生的概率最大为 0.7 (甚至高于多个原因的情况), 最小为 0.14, 进一步反映出不同原因对航空安全事件的不同影响。

d) 当 3 个原因节点 VS_{i1} 、 VS_{71} 、 VS_{n2} 状态都为 F (即 3 个原因都不存在) 时, VE_2 事件发生的概率为 0。

由此可以进一步推理出不同事件与多种原因的概率关系。

4 结束语

基于贝叶斯网络的民航突发事件因果关系分析方法结合了领域本体的语义知识表达与贝叶斯网络的概率推理能力, 通过领域本体与贝叶斯网络转换规则的设计实现了基于领域本体的贝叶斯网络构建, 采用世界民航事故调查跟踪信息报告中专家的评价给出了贝叶斯网络证据节点的概率属性设置, 将 E-IPFP 概率融合算法与消息传递机制引入贝叶斯网络完成了突发事件因果关系节点变量的后验概率值的计算, 在获得不同类别突发

事件因果关系分布的同时, 有效解决了基于领域本体的民航突发事件因果关系分析, 为进一步基于领域本体的因果关系推理奠定了很好的数据基础。未来将针对海量的领域本体 RDF 图与贝叶斯网络转换及其分布式推理进行深入的研究。

参考文献:

[1] CCAR396-R3, 民用航空安全信息管理规定 [S].

[2] 干红华. 基于事件的因果关系可计算化分析研究 [D]. 杭州: 浙江大学, 2003.

[3] 钟军, 禹龙, 田生伟, 等. 基于双层模型的维吾尔语突发事件因果关系抽取 [J]. 自动化学报, 2014, 40 (4): 771-779.

[4] 周文涛, 王红, 王静. 民航应急决策方案语义模型构建方法的研究 [J]. 计算机应用研究, 2013, 30 (1): 195-198.

[5] 李林, 王红, 付宇, 等. 民航突发事件应急案例语义检索方法研究 [J]. 计算机工程与设计, 2011, 32 (3): 1130-1133.

[6] 王红, 杨璇, 王静, 等. 基于本体的民航应急决策知识表达与推理方法研究 [J]. 计算机工程与科学, 2011, 33 (4): 129-133.

[7] 吴贤国, 丁保军, 张立茂, 等. 基于贝叶斯网络的地铁施工风险管理研究 [J]. 中国安全科学学报, 2014, 24 (1): 84.

[8] 谢红薇, 闫婷, 车晋强. 本体驱动的贝叶斯网络模型在医学诊断中的应用 [J]. 太原理工大学学报, 2016, 47 (3): 389-393.

[9] Elster C, Wübbeler G. Bayesian inference using a noninformative prior for linear Gaussian random coefficient regression with in homogeneous within-class variances [J]. Computational Statistics, 2017, 32: 1-19.

[10] Wood F, Willem V D M J, Mansinghka V. A New Approach to Probabilistic Programming Inference [J]. Computer Science, 2015: 1024-1032.

[11] 蔡炳万, 石宇强, 李明辉, 等. 基于本体的贝叶斯网络知识推理研究 [J]. 机械设计与制造, 2016 (1): 84-87.

[12] 张什永. 贝叶斯网不确定性推理研究 [D]. 合肥: 中国科学技术大学, 2010.

[13] Liu Y, Chen S, Li S, et al. Bayes-SWRL: a probabilistic extension of SWRL [C]// Proc of the 9th International Conference on Computational Intelligence and Security. Washington DC: IEEE Computer Society, 2013: 702-706.

[14] <http://safety.caac.gov.cn/index/initpage.act> [EB/OL]. 2017.

[15] 赵建喆, 李凯. 一种改进的多模块贝叶斯网络局部推理算法 [J]. 东北大学学报: 自然科学版, 2015, 36 (9): 1251-1255.